Self Supervised Learning-based (SSL) models should be able to progressively and continuously learn new tasks without forgetting the previous ones, and also, whenever new unlabeled data is available.

Retraining the model with the full dataset (old+new) is impractical, costly, and even impossible when previous data is not available anymore.

Our work is the first application of continual learning to the field of handwritten text recognition, showing its ability to continuously learn from new scripts or languages.

An efficient continual self-supervised learning framework in terms of complexity and memory usage is proposed to address the issue of catastrophic forgetting with no need for prior knowledge at inference time.

Our approach addresses privacy concerns prevalent in various scenarios, where the storage of entire images is impractical, costly, and even impossible when previous data is not available anymore.

Datasets:
Three examples of the images from the different languages/scripts datasets that were used for the experiments: IAM[3], LAM4[4] and HKR5.

Motivation

- Self Supervised Learning-based (SSL) models should be able to progressively and continuously learn new tasks without forgetting the previous ones, and also, whenever new unlabeled data is available.
- Retraining the model with the full dataset (old+new) is impractical, costly, and even impossible when previous data is not available anymore.
- Our work is the first application of continual learning to the field of handwritten text recognition, showing its ability to continuously learn from new scripts or languages.
- An efficient continual self-supervised learning framework in terms of complexity and memory usage is proposed to address the issue of catastrophic forgetting with no need for prior knowledge at inference time.
- Our approach addresses privacy concerns prevalent in various scenarios, where the storage of entire images containing sensitive information may not be feasible.

Datasets and metrics

Datasets:
Three examples of the images from the different languages/scripts datasets that were used for the experiments: IAM[3], LAM4[4] and HKR5.

Metrics:
For the evaluation, we use the Character Error Rate (CER) between the produced text output and the ground truth.

Implementation details:

- **ConfFigure:** Pre-Training Fine-tuning
- **Optimizer:** AdamW Adam
- **Learning rate:** 5e-05 3e-05
- **Weight decay:** 1e-05 5e-05
- **Optimization momentum:** 0.9 0.9
- **Batch size:** 32 32
- **Learning rate schedule:** cosine decay cosine decay
- **Warmup epochs:** 3 3

The obtained CER after the continual pre-training:

1. Evaluation of continually pre-trained models:

   - **Method**
   - **Language**
   - **Dataset**
   - **Tasks**
   - **# Scripts**
   - **# Languages**
   - **# Samples**
   - **Pre-trained**
   - **Fine-tuned**

2. Comparison with state-of-the-art approaches for HTR:

   - **System**
   - **Language**
   - **Dataset**
   - **Tasks**
   - **# Scripts**
   - **# Languages**
   - **# Samples**
   - **Pre-trained**
   - **Fine-tuned**

Conclusion and Futures works

Our proposed CSSL-MHTR approach consists of an encoder-decoder transformer model that includes language/script adapter components and a memory replay strategy for continual self-supervised learning on handwritten text images.

As future work, we plan to extend this approach to recognize full pages instead of segmented lines. Also, we will explore the addition of other document analysis tasks to be learned continually, for instance, layout analysis, name entity recognition and information extraction.

References


